

Stufflebeam, Daniel L.

## Evaluation als Entscheidungshilfe

Wulf, Christoph [Hrsg.]: *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen*. München : R. Piper & Co. Verlag 1972, S. 113-145. - (Erziehung in Wissenschaft und Praxis; 18)



### Quellenangabe/ Reference:

Stufflebeam, Daniel L.: Evaluation als Entscheidungshilfe - In: Wulf, Christoph [Hrsg.]: *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen*. München : R. Piper & Co. Verlag 1972, S. 113-145 - URN: urn:nbn:de:0111-opus-14250 - DOI: 10.25656/01:1425

<https://nbn-resolving.org/urn:nbn:de:0111-opus-14250>

<https://doi.org/10.25656/01:1425>

### Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, veröffentlichen oder widernatürlich nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

### Kontakt / Contact:

peDOCS  
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Digitalisiert

Mitglied der

  
Leibniz-Gemeinschaft

# Evaluation

Beschreibung und Bewertung von Unterricht,  
Curricula und Schulversuchen

Texte

herausgegeben von Christoph Wulf



R. Piper & Co. Verlag  
München

ISBN 3-492-01985-4  
© R. Piper & Co. Verlag, München 1972  
Gesamtherstellung Clausen & Bosse, Leck/Schleswig  
Umschlagentwurf Gerhard M. Hotop  
Printed in Germany

DANIEL L. STUFFLEBEAM

## *Evaluation als Entscheidungshilfe*

In den letzten zweieinhalb Jahren habe ich mit Mitarbeitern aus Schulen, Erziehungsministerien mehrerer Bundesstaaten und dem amerikanischen Erziehungsministerium intensiv an Problemen der Evaluation gearbeitet. In den meisten Fällen galt es, Projekte zu evaluieren, die aus Titel I und Titel III des Elementary and Secondary Education Act von 1965 finanziert wurden. Der vorliegende Beitrag beruht auf diesen Erfahrungen und stellt den Versuch dar, einige meiner Vorstellungen über die Aufgaben der Evaluation im Rahmen der gegenwärtigen innovativen Bildungsprogramme darzustellen.

Der Aufsatz umfaßt zwei Teile. Im ersten Teil soll der gegenwärtige Stand der Evaluation im Bildungswesen dargestellt werden. Es gilt die Aufgaben der Evaluation zu beschreiben und zu zeigen, daß Pädagogen bislang bei dem Versuch, diesen Aufgaben gerecht zu werden, ineffektiv waren. Sodann sollen einige mögliche Gründe für das unzulängliche Ausmaß an Evaluation im Bildungswesen genannt werden. Im zweiten Teil des Beitrags versuche ich dann einige alternative Ansätze der pädagogischen Evaluation zu entwickeln. Es gilt Evaluation zu definieren, vier m. E. für Innovationen im Bildungsbereich besonders wichtige Evaluationsstrategien zu entwickeln und die Struktur von Evaluationsplänen zu erklären.

### *I. Der gegenwärtige Stand der Evaluation im Bildungswesen*

Erziehung wird in immer stärkerem Maße als ein Mittel zur Befriedigung der sozialen, wirtschaftlichen und geistigen Bedürfnisse der Gesellschaft angesehen. Um dieser ständig schwieriger werdenden Rolle gerecht zu werden, müssen Pädagogen sich mit den wichtigen gesellschaftlichen Problemen befassen. Zu ihnen gehören die Fragen der Chancengleichheit der verschiedenen Rassen, der De-facto-Segregation, der Aufstände in den Städten, der Desillusionierung der Jugend und der zahlreichen vorzeitigen

Schulabgänge. Den hier sich zeigenden Tendenzen und ihrer Ausweitung muß im Interesse unserer Gesellschaft entgegengearbeitet werden. Der Erziehung kommt dabei eine wichtige und schwierige Aufgabe zu, zu deren Bewältigung zahlreiche Reformen erfolgen müssen.

### *Voraussetzungen*

Um den Pädagogen die Erfüllung ihrer neuen Aufgaben zu ermöglichen, stellt die Gesellschaft in allen Bildungsbereichen jährlich viele Milliarden Dollar im Rahmen der einzelstaatlichen und bundesstaatlichen Programme und mit Hilfe von Stiftungen zur Verfügung. Ein Beispiel für eine derartige Investition im Bildungswesen sind der Elementary and Secondary Education Act von 1965, das Head Start Program, der Education Professions Act und das Experienced Teacher Fellowship Program. Außerdem entwickelten viele Industriezweige Bildungsinitiativen, so daß es voraussichtlich bald viele von der Industrie finanzierte Bildungsprojekte geben wird. Aufgrund der neuen Aufgaben gibt es im Bildungswesen noch nie dagewesene Möglichkeiten für die Entwicklung innovativer Programme.

Diese Situation hat jedoch auch dazu geführt, die Evaluation neuer Bildungspläne und Bildungsprogramme zu verlangen. Das gilt vor allem für die aus Bundesmitteln finanzierten Programme wie Titel I und Titel III des Elementary and Secondary Education Act. Hier fordert das Gesetz ausdrücklich, daß die finanzierten Projekte mindestens einmal jährlich einen Evaluationsbericht einreichen. Deshalb müssen viele Pädagogen in allen Sektoren des Bildungswesens erstmals eine formale Evaluation durchführen.

Die Vorschrift, solche Evaluationsuntersuchungen durchzuführen, ist sinnvoll und meiner Meinung nach längst fällig gewesen. Geldgeber und Öffentlichkeit haben das Recht, zu erfahren, ob ihre hohen Bildungsausgaben die erwünschten Erfolge erzielen. Noch stärker benötigen die Pädagogen selbst evaluative Informationen, um eine rationale Grundlage für Entscheidungen über alternative Pläne und Verfahren zu haben. Die Forderung nach der Evaluation von Bildungsprogrammen bewirkt jedoch noch nicht ihre Operationalisierung. Pädagogen müssen die Notwendigkeit der Evaluation einsehen und wirksame Evaluationsuntersuchungen durchführen.

### *Die Notwendigkeit besserer Evaluation im Bildungswesen*

Ohne Zweifel sind viele Pädagogen davon überzeugt, daß Bildungsprogramme evaluiert werden müssen. Die Mehrzahl der gegenwärtig vorhan-

denen Evaluationsberichte der Schulen, Erziehungsministerien und Regional Educational Laboratories lassen erkennen, daß Pädagogen viel Zeit, Anstrengung und Geld für die Evaluation ihrer Programme aufwenden. Das hat jedoch noch nicht dazu geführt, eine wirkungsvolle Evaluation durchzuführen. Denn obwohl Pädagogen zahlreiche Evaluationsuntersuchungen gemacht haben, haben ihre Bemühungen nicht dazu beigetragen, die Informationen zu gewinnen, die als Grundlage für Entscheidungen über die evaluierten Programme notwendig sind.

Viele Evaluationsberichte enthalten nur bruchstückhafte Informationen. Obwohl solche Informationen für die Entscheidungsträger wichtig sein können, fehlt ihnen jedoch im allgemeinen das Ausmaß an Zuverlässigkeit, das Entscheidungsträger brauchen, um ihre Entscheidungen zu rechtfertigen, so daß derartige Informationen für wichtige Entscheidungen nur selten nützlich sind. Ein Beispiel dafür ist der erste Jahresbericht für den Titel I des Elementary and Secondary Education Act<sup>1</sup>. Dieser Bericht war sehr wichtig, da er die vielen tausend Projekte des Titel I umfaßte. Sein Wert wurde jedoch dadurch erheblich eingeschränkt, daß er fast keine empirisch gewonnenen Daten enthielt. Statt dessen bot er viele anekdotische Berichte, in denen Projektleiter darlegten, daß ihrer Meinung nach ihr Programm erfolgreich sei; viele von ihnen dachten sogar darüber nach, welches die Gründe für die angeblichen Erfolge sein könnten. Obwohl diese Anekdoten manchmal Fragen anschnitten, die für die Verbesserung des Titel I Program von Bedeutung waren, konnten die Entscheidungsträger im Kongreß, im amerikanischen Erziehungsministerium, in den Erziehungsministerien der Bundesstaaten und in den örtlichen Schulbezirken wichtige Entscheidungen kaum auf solche Beweisstücke gründen.

Beim Titel III des Elementary and Secondary Education Act ist die Situation kaum anders. Die für den Titel III verantwortlichen Beamten im amerikanischen Erziehungsministerium bestimmten die Qualität der Titel-III-Anträge immer aufgrund von 15 Kriterien mit Hilfe einer Fünf-Punkte-Skala<sup>2</sup>. Die Beurteilung im Kriterium, das sich auf die Evaluation bezog, lag in der Regel im negativen Bereich der Skala und war schlechter als bei dreizehn der anderen Kriterien; Ausnahme war das Kriterium, das sich auf die Dissemination bezog. Guba hat überzeugend dargelegt, daß die Evaluationspläne in den Anträgen zu Titel-III-Projekten unzureichend sind<sup>3</sup>. Aufgrund einer Analyse von 32 Anträgen zu Titel-III-Projekten kam Guba zu folgendem Ergebnis: »Es ist fraglich, ob die Ergebnisse dieser Evaluationsuntersuchungen überhaupt nützlich sind. Sie entsprechen wahrscheinlich der unter Pädagogen verbreiteten stereotypen Auffassung von Evaluation als etwas, das von oben gefordert wird und zu dessen Herstellung Zeit und Mühe erforderlich ist, das aber für Handlungsabläufe nur wenig re-

levant ist.«<sup>4</sup> Im Unterschied zu den bereits erwähnten Evaluationsuntersuchungen von Titel I und Titel III enthalten einige andere empirisch gewonnene Daten. So wurden z. B. für den Evaluationsbericht des New York City Higher Horizons Program exakte Forschungsverfahren benutzt (Wrightstone o. J.), um die Leistungen einer Versuchsgruppe, die im Rahmen dieses Programms unterrichtet wurde, mit denen einer Kontrollgruppe zu vergleichen, die in einigen Punkten mit der Versuchsgruppe parallelisiert worden war. Die Ergebnisse dieses fast 300 Seiten umfassenden Berichts waren für die Ergebnisse genauer Evaluationsuntersuchungen typisch. Es gab keine signifikanten Unterschiede. Im deutlichen Unterschied dazu stellte der Bericht jedoch auch fest, daß nach der Überzeugung der an dem Programm beteiligten Lehrer und Schulleiter die Unterschiede so stark waren, daß man das Programm auf keinen Fall wieder aufgeben dürfe.

Obwohl die Evaluationsuntersuchungen der Projekte der Titel I und III sich von der Evaluation des Higher Horizons Program in ihrer Genauigkeit unterschieden, waren sie in einer Hinsicht jedoch gleich. Keine von ihnen half den Entscheidungsträgern, die evaluierten Programme zu verbessern. Obwohl ich nur drei Beispiele für Unzulänglichkeiten bei gegenwärtigen Evaluationsuntersuchungen angeführt habe, sind sie meiner Meinung nach genügend aussagekräftig, um meinen Standpunkt zu veranschaulichen. In vielen Fällen helfen Evaluationsberichte den Entscheidungsträgern nur wenig oder gar nicht, so daß Entscheidungen im Bildungswesen zu treffen intuitiv und riskant bleibt.

### *Probleme der Evaluation im Bildungswesen*

Wie läßt sich diese Situation erklären? Warum können Pädagogen nicht Evaluationsuntersuchungen durchführen, die zugleich nützlich und wissenschaftlich einwandfrei sind? Warum gewinnt man aus Evaluationsuntersuchungen mit Hilfe klassischer Forschungsmethoden nur Informationen, die für Entscheidungen über Bildungsprogramme von begrenztem Wert sind? Warum stehen die Ergebnisse vieler Evaluationsuntersuchungen, die keinen signifikanten Unterschied aufweisen können, in Widerspruch zu den Erfahrungen der an dem Programm Beteiligten?

Man kann diesen Fragen nicht einfach mit dem Argument begegnen, daß die Praxis der Evaluation zu weit hinter den Ansprüchen der Theorie zurückbleibt oder daß die Pädagogen sich nicht genügend bemühen, ihr Programm zu evaluieren. Auch sollte man nicht die evaluativen Äußerungen der beteiligten Personen als unglaublich hinstellen oder behaupten, daß Ergebnisse mit nicht signifikanten Unterschieden typisch sind,

weil in der Pädagogik alle Anstrengungen kaum jemals einen Unterschied bewirken. Nach meiner Meinung besteht der Mangel an angemessenen Evaluationsdaten, weil verschiedene grundlegende Probleme erst gelöst werden müssen, bevor die Evaluationsuntersuchungen sich verbessern lassen. Zu diesen Problemen gehört das Fehlen ausgebildeter Evaluatoren, adäquater Evaluationsinstrumente und Evaluationsverfahren und einer angemessenen Theorie der Evaluation. Nach meiner Auffassung liegt das größte Problem im Fehlen einer für die Evaluation von Bildungsprogrammen geeigneten Konzeptualisierung oder Theorie.

Die konzeptuellen Grundlagen sind für Evaluationsuntersuchungen von grundlegender Bedeutung. Wenn die Konzeptionen falsch sind, dann sind die auf ihnen beruhenden Evaluationsergebnisse auch falsch. Daher ist es wichtig, die Qualität der Konzeptualisierungen zu untersuchen, die den gegenwärtigen Anforderungen an Evaluation zugrunde liegen. Es empfiehlt sich, die Konzeptualisierungen in folgende drei Klassen einzuteilen und jede getrennt zu betrachten:

1. Konzeptionen von der Beschaffenheit der Bildungsprogramme, für die Evaluationsuntersuchungen gebraucht werden, z. B. von den Entscheidungsprozessen und den entsprechenden Informationsbedürfnissen, die die Evaluationsuntersuchungen befriedigen sollen
2. Konzeptionen vom Wesen der Evaluation und ihrer Beziehung zu einzelnen Klassen von Bildungsprogrammen
3. Konzeptionen von der Struktur der Evaluationspläne, die für die Durchführung pädagogischer Evaluationsuntersuchungen gebraucht werden.

### Probleme bei der Bestimmung der Anforderungen für Evaluation im Bildungswesen

Zunächst wollen wir einmal die Probleme untersuchen, die sich bei der Bestimmung der Aufgabe und Funktion pädagogischer Evaluationsuntersuchungen ergeben. Um eine Evaluation machen zu können, muß man natürlich erst wissen, was evaluiert werden soll. Das zu wissen, ist jedoch gegenwärtig eine außerordentlich schwierige Aufgabe. Die augenblicklichen Bemühungen um Evaluation sind aufgrund der neuen pädagogischen Programme und Aktivitäten entstanden. Zu diesen Aktivitäten gehören auch die erst in letzter Zeit für die Pädagogen entstandenen neuen Aufgaben, die neuen Verhältnisse in den verschiedenen Bereichen des Bildungswesens und das Anliegen zahlreicher Institutionen, zu gemeinsamen Bildungsentscheidungen zu kommen. Daher sollte man sich nicht dadurch beirren lassen, daß die bislang für das Bildungswesen gültige Theorie der Evaluation nicht länger ausreichte, um die Informationen, die für die Entwicklung



der neuen Bildungsprogramme notwendig sind, zu gewinnen. Viele neue Bildungsprogramme unterscheiden sich von den bisherigen so sehr, daß unsere Evaluationsuntersuchungen Fragen beantworten müssen, die sich von denen der Vergangenheit stark unterscheiden.

Meiner Meinung nach brauchen wir Konzeptualisierungen, die den Entscheidungsprozessen und Informationsbedürfnissen bei den neuen Bildungsprogrammen gerecht werden. Solche Programme, die zur Verbesserung des Bildungswesens beitragen sollen, sind von zahlreichen unterschiedlichen Entscheidungen abhängig; um diese Entscheidungen zu fällen, werden Informationen benötigt. Evaluatoren, die diese Informationen beschaffen sollen, müssen die relevanten Entscheidungsprozesse und entsprechende Informationsbedürfnisse kennen, bevor sie angemessene Evaluationsuntersuchungen planen können. Sie müssen den Ort, den Schwerpunkt, den Zeitpunkt und die kritische Reflektiertheit der Entscheidungen kennen, die sie vorbereiten sollen. Gegenwärtig gibt es weder eine adäquate Kenntnis der Entscheidungsprozesse und der entsprechenden Informationsbedürfnisse bei Bildungsprogrammen, noch gibt es ein systematisches Programm, diese Kenntnisse zu gewinnen. Kurz gesagt, es gibt keine angemessenen Konzeptualisierungen der Entscheidungen und der entsprechenden Informationsbedürfnisse; es gibt auch keine Vorstellungen über Programme, mit deren Hilfe sie gewonnen werden können.

### Probleme bei der Definition von Evaluation im Bildungswesen

Als nächstes wollen wir Fragen nach der Bedeutung der Evaluation im Bildungswesen erörtern. Im allgemeinen haben Pädagogen es als die Aufgabe der Evaluation angesehen, das Ausmaß zu bestimmen, in dem Lernziele erreicht worden sind. Der erste Schritt zur Operationalisierung dieser Definition besteht darin, Programmziele in Verhaltensbegriffen zu formulieren. Um eine Beziehung zwischen Ergebnissen und Zielen herzustellen, muß man sodann Kriterien definieren und operationalisieren. Zur Operationalisierung dieser Kriterien gehört auch die Spezifikation der Instrumente, mit deren Hilfe die Ergebnisse und die Normen gemessen werden, die zur Bewertung der Ergebnisse dienen sollen.

Normen sind entweder absolut oder relativ. Eine absolute Norm könnte z. B. in einer bestimmten Punktzahl bestehen, die alle Schüler als Mindestdurchschnitt in einem ausgewählten Leistungstest erreichen sollen. Eine relative Norm könnte z. B. dadurch gebildet werden, daß eine Schülergruppe, die mit einem neuen Programm arbeitet, in einem ausgewählten Leistungstest im Durchschnitt höhere Punktzahlen erreichen soll als eine entsprechende Schülergruppe, die mit einem herkömmlichen Programm ar-

beitet. Ungeachtet der verwendeten Evaluationsnorm werden bei einer solchen Untersuchung die Daten erst nach einem vollständigen Ablauf des Programms analysiert, um zu bestimmen, in welchem Ausmaß die Ziele erreicht worden sind.

Evaluationsuntersuchungen, die nach diesem Modell durchgeführt werden, liefern nach Ablauf des Programms Daten über seine Gesamtwirkung und helfen Entscheidungen über das Programm zu fällen. Sie unterstützen aber den Pädagogen nicht bei der Anfangsplanung und der Realisierung der Programme. Solche Evaluationsuntersuchungen bieten daher nur eine ungenügende Hilfe für die Lösung der Probleme der Pädagogen, die innovative Programme planen und durchführen müssen.

Die Unzulänglichkeit der vorhandenen Evaluationskonzepte wird durch den folgenden Auszug aus den Ausführungen über die Evaluationsuntersuchungen des Titel I belegt, die eine Gruppe New Yorker Bürger vor einer Kommission des Kongresses machte: »Wir bitten um eine Verbesserung der gesetzlichen Bestimmungen über die Evaluation von Titel-I-Projekten, damit die Ergebnisse der Evaluation besser genutzt werden. Das Gesetz bestimmt lediglich, daß Evaluationsuntersuchungen gemacht werden müssen, nicht jedoch, daß ihre Ergebnisse für eine zukünftige Planung verwendet werden sollen. In New York City wurden in diesem Jahr Projekte wiederholt durchgeführt, ohne daß die Evaluationsergebnisse des vergangenen Jahres vorlagen. Um Evaluationsuntersuchungen wirksamer zu machen, sollten sie auch Alternativen und Empfehlungen des Evaluators enthalten. Was bislang nur eine kostspielige, für die Programmentwicklung sekundäre Aktivität war, sollte eine echte Aufgabe des Evaluators werden, damit den örtlichen Schulverwaltungen geholfen wird, ihre Entscheidungen auf Erfahrungen und empirische Daten zu gründen. Die amerikanische Wirtschaft wäre auch nicht funktionsfähig, wenn nicht ihre Berater, sobald sie die Wirksamkeit bestimmter Programme untersucht haben, das Management mit Alternativen versorgen würden.«<sup>5</sup>

Nach dieser Auffassung sind die aufgrund der gegenwärtigen Evaluationsprogramme verfaßten Berichte weder spezifisch noch rechtzeitig genug verfügbar, um Bildungsprogramme beeinflussen zu können. Evaluationsuntersuchungen, die diesen beiden Kriterien jedoch nicht gerecht werden, sind nur von geringem Nutzen.

### Probleme der Planung von Evaluation im Bildungswesen

Schließlich sollen Probleme der Methodologie der Evaluation erörtert werden. Wenn die gegenwärtigen Konzeptionen der Evaluation sich nicht für die Evaluation heutiger pädagogischer Aktivitäten eignen, können auch

die entsprechenden Evaluationspläne nicht angemessen sein. Die bestehenden Verfahren der Evaluation wurden entwickelt, um die Aufgaben der Evaluation so zu erfüllen, wie sie in der Vergangenheit bestimmt worden waren.

Die Unzulänglichkeit der vorhandenen Evaluationsmethodologie wird deutlich, wenn man untersucht, welche Pläne Pädagogen für die Evaluation ihrer Programme verwenden. Wenn sie überhaupt einen Evaluationsplan entwickelt haben, ist es im allgemeinen ein experimenteller Versuchsplan. Sein wichtigstes Ziel besteht darin, Daten so zu erheben, daß sie eine innere Validität (*internal validity*) haben. Dazu müssen einige Bedingungen erfüllt werden. Die Einheiten des Bildungsprogramms, die untersucht werden sollen, müssen nach dem Zufallstichproben-Verfahren den Versuchs- und den Kontrollbedingungen zugeteilt werden. Eine Reihe von Schülern kann z. B. nach dem Zufallstichproben-Verfahren in zwei Gruppen geteilt werden; eine davon arbeitet mit dem neuen Programm, die andere mit dem herkömmlichen Programm der Schule. Sodann müssen die Versuchs- und Kontrollbedingungen geschaffen werden, die während des ganzen Versuchs konstant gehalten werden und der Anfangsdefinition der Bedingungen entsprechen müssen. Die Bedingungen des neuen Programms dürfen im Verlauf des Untersuchungsprozesses nicht verändert werden, da man sonst nicht weiß, was wirklich evaluiert worden ist.

Alle an dem Experiment teilnehmenden Schüler müssen in gleichem Ausmaß den Bedingungen ihrer Gruppe ausgesetzt werden; es muß darauf geachtet werden, daß die Schüler der Versuchsgruppe von denen der Kontrollgruppe deutlich getrennt sind. Denn wenn eine Kontaktaufnahme zwischen den Gruppen stattfindet, kann man nach Abschluß des Projekts nicht mehr feststellen, welche Ergebnisse durch welche Bedingungen verursacht worden sind. Daher muß man bis nach Abschluß des Experiments der Versuchung widerstehen, die erfolgreichen Aktivitäten der Versuchs- oder der Kontrollgruppe den an diesem Versuch in einer der beiden Gruppen teilnehmenden Schülern zugute kommen zu lassen, selbst wenn die Aktivitäten in einer Gruppe der Schüler sehr unzureichend sind.

Schließlich muß ein Instrument, das für ein bestimmtes Kriterium valide und reliabel ist, nach einer gewissen Zeitspanne – im allgemeinen nach einem ganzen Programmablauf – Versuchspersonen aus beiden Gruppen des Experiments erneut vorgelegt werden. Wenn alle diese Bedingungen erfüllt würden, könnte man mit Hilfe statistischer Verfahren und Entscheidungsregeln unzweideutig bestimmen, ob es zwischen den Versuchs- und den Kontrollgruppen im Hinblick auf die interessierenden Variablen signifikante Unterschiede gegeben hat oder nicht.

Auf den ersten Blick scheint sich die Anwendung eines experimentellen

Versuchsplans auf Evaluationsprobleme zu empfehlen, da bisher experimentelle Forschung und Evaluation dazu verwendet worden sind, Hypothesen über die Auswirkungen bestimmter Programme zu überprüfen. In dieser Aufgabenbestimmung sind jedoch vier wichtige Probleme enthalten:

*Erstens gerät die Anwendung eines experimentellen Versuchsplans auf Evaluationsprobleme mit der Aufgabe der Evaluation, zur kontinuierlichen Verbesserung eines Bildungsprogramms beizutragen, in Konflikt.* Ein experimenteller Versuchsplan verhindert die Modifikation der Versuchsbedingungen eher, als daß er sie fördert, weil die Versuchs- und Kontrollbedingungen im Verlauf des Versuchs nicht verändert werden dürfen, ohne daß die Daten über die Unterschiede zwischen den Versuchs- und den Kontrollgruppen verfälscht werden. Somit richten sich die Versuchs- und die Kontrollbedingungen nach dem Evaluationsplan anstatt umgekehrt; der Versuchsplan verhindert also eher Veränderungen in den Bedingungen, als daß er sie fördert.

Man kann nicht erwarten, daß die Leiter von Innovationsprojekten sich den Bedingungen eines experimentellen Versuchsplans unterwerfen. Sie können die Entwicklung ihres Projekts nicht im Anfangsstadium lassen, nur um am Jahresende Evaluationsdaten mit innerer Validität zu haben. Die Projektleiter müssen vielmehr alle erhältlichen Daten verwenden, um den Projektplan und seine Implementation kontinuierlich zu verbessern oder ihn in einigen Fällen von Grund auf zu verändern. Deshalb braucht man Konzeptionen von Evaluationsprogrammen, die eine dynamische Entwicklung der Bildungsprogramme fördern und nicht hemmen.

*Eine zweite Unzulänglichkeit des experimentellen Versuchsplans besteht darin, daß er zwar nach Abschluß eines Projekts Daten für Entscheidungen zur Verfügung stellt, daß er jedoch für Entscheidungen während der Planung und Implementation eines Projekts fast nutzlos ist.* Er liefert nach Abschluß des Versuchs Daten über die relative Wirkung der Programme in den Versuchs- und Kontrollgruppen. Solche Daten sind jedoch weder genügend spezifisch und umfassend, noch stehen sie zur rechten Zeit zur Verfügung, um dem Entscheidungsträger bei der Bestimmung der Projektziele und des Projektplans oder bei der Modifikation seiner Implementation behilflich zu sein. Bestenfalls zeigen experimentelle Versuchspläne hinterher, ob ein Projekt seine Ziele erreicht hat. Dann ist es jedoch zu spät, Entscheidungen über Pläne und Verfahren zu treffen, die bereits weitgehend den Erfolg oder Mißerfolg eines Projekts bestimmt haben.

*Guba hat auf ein drittes mit dem experimentellen Versuchsplan verbundenes Problem hingewiesen: Dieser Plan eignet sich eher für die antiseptischen Bedingungen des Laboratoriums als für die septischen Bedingungen*

der Schule<sup>6</sup>. Die potentiellen Störvariablen (confounding variables) müssen entweder kontrolliert oder durch Randomisierung eliminiert werden, wenn die Ergebnisse eine inhaltliche Validität haben sollen. Im Bereich der Erziehung gelingt dies jedoch fast nie. Untersuchen wir z. B. das folgende Zitat aus einem Evaluationsbericht von Julian Stanley (1966):

»Selbst wenn das Programm einen hohen positiven Einfluß auf die Berufslaufbahn einer Person hat, tritt dieser vielleicht nur langsam in Erscheinung und kann so mit anderen Einflüssen verbunden sein, daß ihn selbst die betreffende Person nicht deutlich erkennen kann. Dennoch müssen wir alle Daten zu der Entscheidung darüber heranziehen, ob sich die wiederholte Benutzung von bestimmten Bildungsprogrammen empfiehlt, bzw. welche ihrer Teile modifiziert werden müssen, um ihre Wirkung zu verbessern. Bei für einen experimentellen Versuchsplan optimalen Bedingungen müßten wir das Programm als kontrolliertes Experiment mit einer parallelierten Kontrollgruppe, die nicht an dem Sommerkurs teilnimmt, durchführen, müßten dann beide Gruppen über mehrere Jahre hinweg weiter verfolgen, um die Unterschiede zwischen ihnen bestimmen zu können. Wenn die Auswahl der Teilnehmer für den Sommerkurs rechtzeitig beginnt und die Gruppe der Bewerber so groß ist, daß aus ihr zwei Gruppen mit genügend hohem Niveau gebildet werden können, kann der experimentelle Versuchsplan verwendet werden. Dennoch können auch in diesem Fall die Reaktionen der abgewiesenen Bewerber und die fehlende Möglichkeit, ihre Aktivitäten während des Sommerkurses zu kontrollieren, eine unerwünschte Auswirkung auf das Ergebnis des Experiments haben. Die Teilnahme an einem angesehenen Programm bestätigt zu bekommen, dürfte bereits eine wirkungsvolle Hilfe sein. . . . Unser wichtigstes Mittel für die Evaluation des Programms waren die Berichte der Projektmitarbeiter und vor allem der Teilnehmer.«

In diesen Ausführungen hat Stanley viele Gründe dafür genannt, warum ein experimenteller Versuchsplan sich nicht für die Evaluation im Bildungswesen eignet. In zahlreichen innovativen Programmen gibt es zu viele Störfaktoren, die sich nicht wirksam kontrollieren lassen.

Die von Stanley z. B. erwähnten Störfaktoren weisen auf ein viertes Problem bei der Anwendung des experimentellen Versuchsplans hin. *Während die innere Validität (internal validity) durch die Kontrolle äußerer Variablen (extraneous variables) erreicht werden kann, wird dieser Gewinn allerdings auf Kosten der äußeren Validität (external validity) erreicht.* Wenn die äußeren Variablen streng kontrolliert werden, sind die Ergebnisse so lange zuverlässig, wie eine Innovation unter kontrollierten Bedingungen stattfindet. Die Ergebnisse solcher Untersuchungen können jedoch nicht auf die Schulwirklichkeit übertragen werden, in der die äußeren Variablen sich nicht kontrollieren lassen. Deshalb gilt es in Erfahrung zu bringen, wie pädagogische Innovationen sich in der Schulwirklichkeit bewähren.

Bislang habe ich in diesem Beitrag den gegenwärtigen Stand der Be-

mühungen um Evaluation im Bildungswesen darzustellen versucht. Am Anfang meiner Ausführungen legte ich dar, daß Pädagogen mit vielen neuen und unterschiedlichen Forderungen nach Evaluation konfrontiert werden. Sodann versuchte ich nachzuweisen, daß die Anstrengungen der Pädagogen, diesen Ansprüchen zu genügen, bislang nicht ausreichen. Schließlich habe ich auf drei Unzulänglichkeiten hingewiesen, die Pädagogen daran hinderten, ergiebige Evaluationsuntersuchungen zu machen:

- (1) Fehlende Kenntnis der Entscheidungsprozesse und Informationen, die bei der Entwicklung innovativer Bildungsprogramme benötigt werden,
- (2) Fehlen einer Definition der Evaluation, die den sich abzeichnenden Forderungen nach Evaluation im Bildungswesen gerecht wird,
- (3) Fehlen angemessener Evaluationspläne

## *II. Das Wesen der Evaluation*

Da dies ein Arbeitsbericht ist, sollte ich nicht länger die gegenwärtigen Bedürfnisse und Probleme der Evaluation behandeln. Man kann meine Ausführungen überprüfen, modifizieren oder ablehnen. Sobald man zu einem Konsens darüber gekommen ist, welches die wirklichen Probleme der Evaluation sind, könnte man relevante Lösungen entwickeln. Hier sollte ich einige Vorstellungen zur Lösung der gegenwärtigen Schwierigkeiten vortragen. Daher werde ich im Verlauf dieses Beitrags einige alternative Konzepte der pädagogischen Evaluation entwickeln.

Der folgende Teil dieses Beitrags gliedert sich in vier Abschnitte. Im ersten soll Evaluation ganz allgemein definiert werden. Im zweiten werden neuere innovative Programme analysiert und die Arten der Entscheidungen identifiziert, für die in diesen Programmen Evaluationsuntersuchungen benötigt werden. Der dritte Teil enthält einen Überblick über vier Strategien zur Evaluation von Bildungsprogrammen. Der Beitrag schließt im vierten Teil mit dem Versuch, die Struktur von Evaluationsplänen zu entwickeln.

### *Merkmale der Evaluation*

#### *Eine rationale Begründung (Rationale)*

Wenn Entscheidungsträger ihre Möglichkeiten maximal ausnutzen wollen, müssen sie vernünftige Entscheidungen über vorliegende Alternativen treffen. Dazu müssen sie jedoch zunächst wissen, welche Alternativen ihnen zur Verfügung stehen. Sie müssen außerdem in der Lage sein, begründete Urteile über den relativen Wert der Alternativen abzugeben. Dies erfor-

dert jedoch relevante Informationen. Die Entscheidungsträger sollten daher über wirksame Mittel verfügen, mit deren Hilfe evaluative Informationen gewonnen werden können. Anderenfalls dürften ihre Entscheidungen von vielen unerwünschten Elementen abhängen. Günstigstenfalls sind die Urteile lediglich von Sympathien, Vorurteilen und Interessen abhängig. Häufig gibt es dabei eine Tendenz, persönliche Erfahrungen, Gerüchte und die Ansicht einer Autorität überzubewerten; so werden gewiß zuviel Entscheidungen getroffen, ohne daß die möglichen Alternativen bekannt sind.

Die Qualität von Programmen hängt von der Qualität der Entscheidungen in den Programmen und über die Programme ab; die Qualität der Entscheidungen wird durch die Fähigkeit der Entscheidungsträger bestimmt, die Alternativen zu identifizieren, die in Entscheidungssituationen auftreten, und entsprechend vernünftige Urteile über sie zu fällen; vernünftige Urteile bedürfen valider und reliabler Informationen über die Alternativen; um solche Informationen zu erhalten und den Entscheidungsträgern zur Verfügung zu stellen, braucht man systematische Verfahren. Die Prozesse, mit deren Hilfe die für die Entscheidungen erforderlichen Informationen gewonnen werden, müssen Teil des Evaluationskonzepts sein. Auf diesen Ausführungen aufbauend, möchte ich eine Definition von Evaluation entwickeln.

### Definition der Evaluation

Im allgemeinen bedeutet Evaluation die Gewinnung von Informationen durch formale Mittel wie Kriterien, Messungen und statistische Verfahren mit dem Ziel, eine rationale Grundlage für das Fällen von Urteilen in Entscheidungssituationen zu erhalten. Zur Erläuterung dieser Definition sollen einige zentrale Begriffe erklärt werden:

Eine Entscheidung ist eine Wahl zwischen Alternativen.

Eine Entscheidungssituation besteht aus einer Reihe von Alternativen.

Ein Urteil fällen bedeutet, die Alternativen zu bewerten.

Ein Kriterium ist ein Maßstab, aufgrund dessen die Alternativen bewertet werden; im Idealfall umfaßt ein Kriterium die Spezifikation von Variablen, Messungen und Normen für die Beurteilung des Untersuchungsgegenstandes.

Statistik ist die Wissenschaft von der Analyse und Interpretation einer Reihe von Meßwerten.

Unter Messung wird die Übertragung von Zahlen auf Einheiten aufgrund bestimmter Regeln verstanden; nach solchen Regeln erfolgt im allgemeinen die Spezifikation der Stichprobenelemente, der Meßverfahren und der Bedingungen für die Durchführung und Beurteilung der Meßverfahren.

Vereinfacht gesagt, ist Evaluation also die Wissenschaft, mit deren Hilfe Informationen für Entscheidungsprozesse zur Verfügung gestellt werden.

Zur Methodologie der Evaluation gehören vier Funktionen: *Sammlung*, *Organisation*, *Analyse* und *Bericht* von Informationen. Zu den Kriterien für die Einschätzung der Angemessenheit der Evaluation gehören *Validität* (Ist die Information diejenige, die der Entscheidungsträger braucht?), *Reliabilität* (Ist die Information reproduzierbar?), *Rechtzeitigkeit* (Steht die Information für den Entscheidungsträger rechtzeitig zur Verfügung?), *Verfügbarkeit* (Erreicht die Information alle Entscheidungsträger, die sie brauchen?) und *Zuverlässigkeit* (Vertrauen die Entscheidungsträger der Information?).

### Evaluation außerhalb des Bildungswesens

Das bisher entwickelte Evaluationskonzept ist sehr allgemein gefaßt, da das Bewerten von Alternativen in allen Bereichen des menschlichen Lebens üblich ist und da Menschen immer bestrebt sind, rational vertretbare Grundlagen für ihre Urteile zu erhalten. Es lassen sich jedoch zahlreiche Arten der Evaluation, die alle die Bedingungen der genannten Definition erfüllen, voneinander unterscheiden. So ist z. B. auch für Marktforschung, Kosten-Nutzen-Analyse (cost-benefit analysis), experimentelle Versuchsplanung, objektive Testverfahren, militärwissenschaftliche Forschung, Planungsforschung, Program Evaluation and Review Technique (PERT), Planning Programming and Budgeting System (PPBS), Qualitätskontrolle und Systemanalyse die erwähnte allgemeine Definition der Evaluation gültig.

Für jedes dieser Forschungsvorgehen ist die Anwendung systematischer Verfahren zur Bewertung von Alternativen in Entscheidungssituationen charakteristisch. Diese verschiedenen Arten der Evaluation lassen sich nach Entscheidungssituationen, Entscheidungsbedingungen, Art der verwendeten Instrumente und Verfahren, Ausmaß der Präzision bei der Sammlung und Analyse von Informationen und den methodischen Fähigkeiten der Evaluatoren und ihrer Adressaten unterscheiden. Diese inhaltlichen und methodischen Unterschiede erklären wahrscheinlich, warum die verschiedenen Formen der Evaluation unterschiedliche Namen haben. Das wird z. B. auch aus folgenden Ausführungen Quades (1967,4) deutlich: »Evaluationsuntersuchungen, die Entscheidungsträgern bei der Wahl zwischen Systemen und bei der Ermittlung ihrer Effektivität im Hinblick auf ihre Ziele oder bei der Entwicklung eines Bezugsrahmens für ihre Erforschung helfen sollen, können selbstverständlich Systemanalysen genannt werden.«

Obwohl Quade behauptet, daß Systemanalyse eine Form der Evaluation



ist, erkennt er zugleich auch, daß die Bezeichnung Systemanalyse wegen der spezifischen Beschaffenheit dieser Art der Evaluation gewählt worden ist.

Betrachtet man die Entstehung der genannten Formen der Evaluation, wird deutlich, daß alle für spezifische Anwendungsbereiche entwickelt worden sind. Program Evaluation and Review Technique (PERT) wurde entwickelt, um dem Militär bei der Entscheidung über die Entwicklung komplexer Waffensysteme zu helfen. Systemanalyse entstand, um dem Militär die Entscheidung über die Entwicklung und Durchführung militärischer Operationen zu erleichtern. Objektive Testverfahren werden beim Militär vor allem zur Auswahl für den Wehrdienst eingesetzt.

Diese Formen der Evaluation wurden rasch entwickelt, da ein großes Bedürfnis nach begründbaren Entscheidungen bestand; sie entsprechen daher auch der Art der erforderlichen Entscheidungen und den Entscheidungsbedingungen. Neue Ansätze der Evaluation wurden entwickelt, weil die bestehenden für die Entscheidungsprozesse nicht genügend Informationen liefern und einmal getroffene falsche Entscheidungen ernsthafte Konsequenzen haben konnten. Militärische Entscheidungen können den Ausgang eines Krieges beeinflussen; also wurden entsprechende Verfahren der militärwissenschaftlichen Forschung, Systemanalyse usw. entwickelt. Wirtschaftliche Entscheidungen können zum Gewinn, Verlust oder Bankrott von Tausenden von Aktionären führen; also wurde die Kosten-Nutzen-Analyse entwickelt.

### *Evaluation im Bildungswesen*

Bislang hatten Entscheidungen im Bildungswesen weniger deutliche Auswirkungen als Entscheidungen in Wirtschaft, Landwirtschaft und Militär. Deshalb fanden auch im Bildungswesen geringere Anstrengungen statt, hochspezialisierte Formen der Evaluation zu entwickeln, um die zahlreichen unterschiedlichen Bildungsentscheidungen zu unterstützen. Die meisten Pädagogen hatten erhebliche Schwierigkeiten, die wichtigsten pädagogischen Entscheidungssituationen zu identifizieren, die im Rahmen der Evaluation besonderer Behandlung bedürfen. Man darf daraus jedoch nicht schließen, daß es in der Pädagogik bislang keine Evaluationsverfahren gegeben hat. Standardisierte Tests wurden bei Entscheidungen über Hochschulzulassungen zur Hilfe herangezogen; sie dienten als Grundlage zur Notengebung, Einstufung der Schüler in das Curriculum und Vergabe von Diplomen. Die Buros Mental Measurement Yearbooks (1965) wurden herausgegeben, um den Pädagogen bei der Auswahl und bei der Anwendung von Tests zu helfen. Kürzlich wurde der Educational Product Information

Exchange (EPIE) eingerichtet<sup>7</sup>, um Pädagogen bei der Auswahl alternativer Unterrichtsmaterialien behilflich zu sein. Abgesehen davon wurden im Bildungswesen bislang jedoch keine speziellen Verfahren entwickelt, die bei Entscheidungen über Bildungsprogramme behilflich sein könnten.

Im Bildungswesen war es durchaus üblich, auch andere Bereiche zu beachten, in denen ähnliche Probleme in Angriff genommen und gelöst wurden. Deshalb haben auch die Pädagogen den experimentellen Versuchsplan als einen Evaluationsplan adaptiert. Dabei wird allerdings ein Verfahren, das zunächst den Bauern bei der Unterscheidung zwischen verschiedenen Arten von Düngemitteln und Saaten helfen sollte, im Erziehungswesen benutzt, um eine Auswahl zwischen alternativen Bildungsinnovationen zu treffen. Offensichtlich ist die Ähnlichkeit zwischen pädagogischen Innovationen und Düngemitteln jedoch außerordentlich gering.

In letzter Zeit hat man die Program Evaluation and Review Technique (PERT), die Systemanalyse und das Planning Programming and Budgeting System (PPBS) im Erziehungswesen angewandt. Obwohl ausgewählte Verfahren aus anderen Bereichen den Pädagogen helfen können, Zeit und Mühen zu sparen, möchte ich aber auch davor warnen, Verfahren aus anderen Bereichen unkritisch zu übernehmen. Anderenfalls könnte es zu einer unangemessenen Anwendung solcher Verfahren auf pädagogische Probleme kommen. Meiner Ansicht nach ist die Anwendung des experimentellen Versuchsplans zur Evaluation innovativer Programme ein Beispiel für die unkritische Übernahme in anderem Zusammenhang entwickelter Verfahren. Die Verwendung des experimentellen Versuchsplans in diesem Kontext hat Pädagogen viel Zeit und Anstrengung gekostet, ohne ihnen bei den Entscheidungsprozessen sehr geholfen zu haben.

Wie bereits dargelegt, braucht meiner Meinung nach das Bildungswesen eine neue Konzeptualisierung, um eine Theorie und Methodologie der Evaluation entwickeln zu können, die für die pädagogischen Probleme relevant ist. Bislang habe ich nur eine allgemeine Begründung und Definition von Evaluation gegeben; im weiteren möchte ich eine rationale Begründung und Definition für Evaluation im Bildungswesen entwickeln.

### **Eine rationale Begründung der Evaluation im Bildungswesen**

Die Programme der Titel I und III des Elementary and Secondary Education Act von 1965 bilden für die Entwicklung einer rationalen Begründung einer pädagogischen Evaluation einen komplexen Kontext. Fast alle Schulbezirke sind an einem oder an beiden Programmen beteiligt. Die Ziele dieser Programme bestehen darin, schulische Leistungen, Erfahrungen und Möglichkeiten sozial benachteiligter Schüler zu verbessern und das Aus-

maß und die Qualität der Innovationen in zahlreichen Bildungsinstitutionen zu erhöhen. Beide Programme finden in allen Teilen der USA Anwendung und sind entsprechend konzipiert. Sie werden auf der Ebene der Bundesstaaten koordiniert, kontrolliert und in den örtlichen Schulbezirken implementiert. Insgesamt stellen beide Programme den örtlichen Schulbezirken jährlich mehr als eine Milliarde Dollar zur Verfügung.

Abbildung 1 stellt eine Konzeptualisierung des Prozesses und der Funktion der Evaluation für Entscheidungsabläufe dar, wie sie in den bundesstaatlichen Programmen bestehen könnten. Eine Reihe von Kontrollschleifen veranschaulicht die Beziehungen zwischen den örtlichen, einzel- und bundesstaatlichen Evaluationsuntersuchungen, denen die Projekte der beiden Programme unterzogen werden. Die Schleife an der rechten Seite veranschaulicht örtliche, die mittlere einzelstaatliche und die linke bundesstaatliche Aktivitäten. Jede Schleife enthält eine Reihe von Blöcken, die die wichtigsten Evaluationsfunktionen repräsentieren.

Block 1 stellt das Bildungsprogramm eines Schulbezirks dar. Es bildet den Kontext, aus dem die Bedürfnisse nach pädagogischen Innovationen entstehen und in dem die Reformen schließlich realisiert werden müssen. Es enthält die *Inputs* des Systems, d. h. Schule, Curriculum, Lehrkörper, Organisation, Politik, Finanzen, schulische Anlagen, Beziehungen zwischen Schule und Gemeinde, und die *Outputs* des Systems, d. h. das kognitive, psychische, physische und soziale Befinden der Schüler und späteren Erwachsenen.

Das erste Segment des Umfangs rechts von Block 1 veranschaulicht die Informationssammlung. Sie findet auf der Ebene der örtlichen Schulbezirke als systematische Sammlung aller Informationen statt, die für spätere Entscheidungen auf der örtlichen, einzelstaatlichen und bundesstaatlichen Ebene benötigt werden.

Block 2 repräsentiert die Organisation der Informationen. Dabei werden die Informationen nach vorher bestimmten Kategorien kodiert, bearbeitet, systematisch gespeichert und im Bedarfsfall abgerufen.

In Block 3 werden die in Block 2 organisierten Informationen unter dem Aspekt der Vorbereitung von Entscheidungsprozessen auf örtlicher, einzelstaatlicher und bundesstaatlicher Ebene analysiert und den örtlichen und einzelstaatlichen Entscheidungsträgern berichtet.

Block 4 stellt die Programmentscheidungen dar, die auf örtlicher Ebene getroffen werden. Zu den örtlichen Entscheidungsträgern, denen die Ergebnisse der Evaluation zur Verfügung gestellt werden, gehören das Board of Education, die Schulverwaltung, der Projektleiter, die Lehrer und der Schulleiter.

Die Entscheidungen, die in Block 4 fallen, werden in Block 5 durchge-

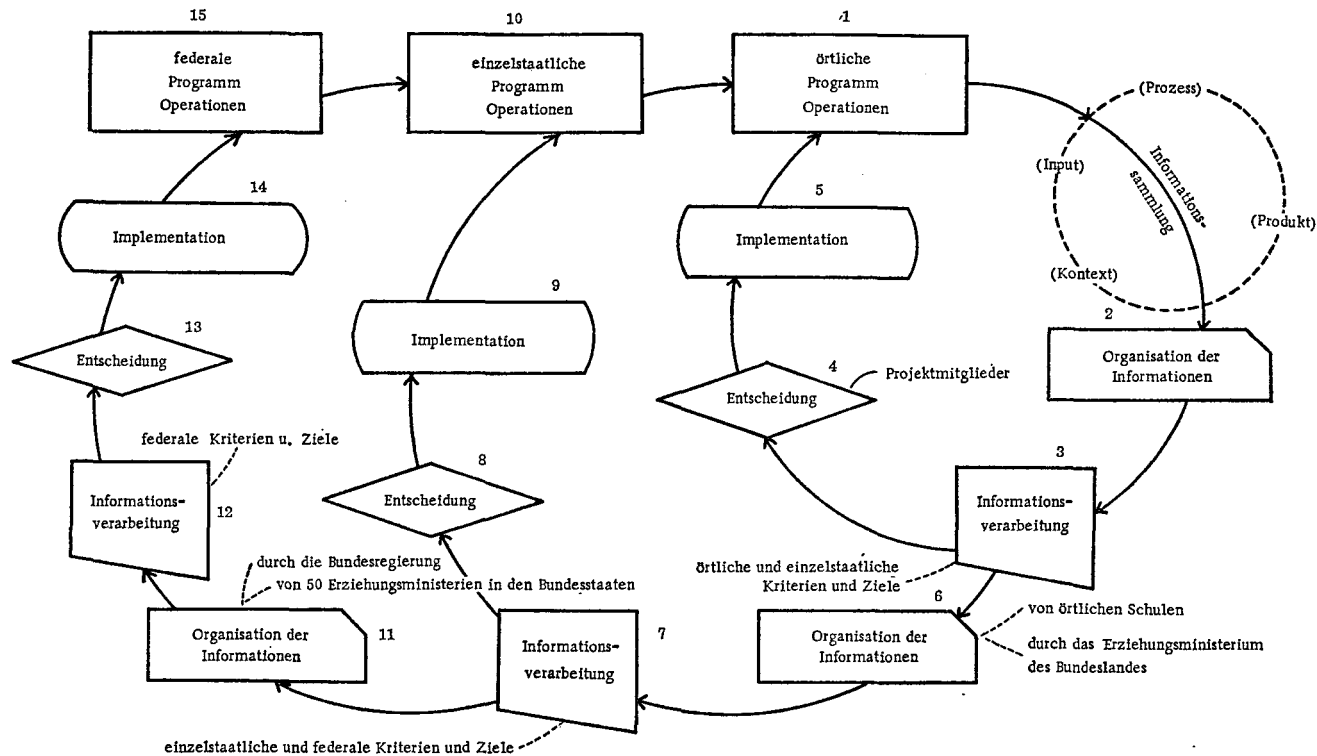


Abb. 1: Evaluation in vom Bund unterstützten Bildungsprogrammen (Stufflebeam 1967)

führt. So beginnt der Zyklus mit zahlreichen Modifikationen des Schulprogramms in Block 1 wieder von neuem.

In Block 3 wird dargestellt – um darauf zurückzukommen –, daß Evaluationsberichte für staatliche Erziehungsministerien jährlich von allen örtlichen Schulbezirken angefertigt werden sollen. In Block 6 würde das Erziehungsministerium eines Bundesstaates dann diese Berichte nach der Art der Projekte organisieren und zu den Informationen über ähnliche Projekte in Beziehung setzen. Diese Informationen würden dann in Block 7 analysiert werden, um die Stärken und Schwächen des Programms im ganzen Bundesstaat zu bestimmen. Die für das Programm in einem Bundesstaat verantwortlichen Beamten würden diese Informationen dazu nutzen, um die pädagogischen Bedürfnisse und Probleme in diesem Staat abzuschätzen, um dann Entscheidungen über Programmschwerpunkte und über die Kontrolle in Block 8 zu treffen. Entscheidungen, die in Block 8 gefällt werden, würden in Block 9 ausgeführt und würden wiederum das Bildungsprogramm des Einzelstaates in Block 10 berühren und den Zyklus in Block 1 wieder von neuem anfangen lassen.

In Block 7 würden jährlich Evaluationsberichte von 50 Staaten an die verantwortliche Bundesinstitution gesandt werden. Die Organisation der Informationen erfolgt dann in Block 11, so daß die wichtigsten Programmansätze aus allen Teilen des Landes in Block 12 auf der bundesstaatlichen Ebene überprüft und analysiert und die Berichte für den Associate Commissioner, der für den Elementary and Secondary Education Act verantwortlich ist, für den Minister, den Präsidenten und den Kongreß vorbereitet werden können. Entscheidungen über Programmschwerpunkte und Finanzen würden auf bundesstaatlicher Ebene in Block 13 gefällt; die Implementation solcher Entscheidungen in Block 14 oder das bundesstaatliche Programm in Block 15 berühren das einzelstaatliche Programm in Block 10 und die örtlichen Projekte der Schulen in Block 1. Somit würde der Zyklus von neuem beginnen.

Zusammengefaßt stellt die Abbildung 1 folgenden Prozeß dar:

- (a) Die Informationen auf bundesstaatlicher, einzelstaatlicher und örtlicher Ebene werden weitgehend in den örtlichen Schulbezirken gesammelt.
- (b) Diese Informationen stellen die Grundlage für bundesstaatliche, einzelstaatliche und örtliche Entscheidungen dar, die schließlich das Handeln in den örtlichen Schulbezirken beeinflussen.
- (c) Evaluationspläne müssen auf der bundesstaatlichen, einzelstaatlichen und örtlichen Ebene entwickelt, verbreitet und koordiniert werden, wenn die Informationen angemessen sind, die die Schulen für die Unterstützung des Entscheidungsprozesses auf allen drei Ebenen zur Verfügung stellen.

Um ein angemessenes Evaluationssystem für Programme wie Titel I und Titel III zu entwickeln, braucht man zunächst einige Kenntnisse der gegebenen Entscheidungssituationen. Die Kenntnis dieser Entscheidungssituationen sollte die Beantwortung einer Reihe von Fragen ermöglichen. Erstens sollte man die Stelle, bzw. *Ebene* der Entscheidungen identifizieren, auf der die Autorität und Verantwortung für die Entscheidungen liegt, d. h. man muß bestimmen, ob die Entscheidungen auf der Ebene der Schulen, der Einzelstaaten oder des Bundes erfolgt. Zweitens sollte man den *Schwerpunkt* (focus) der Entscheidungen bestimmen, also z. B. die Frage, inwieweit sich die Entscheidungen auf die Ziele der Forschung, Entwicklung, Lehrerbildung und Implementation beziehen. Drittens muß man den *Inhalt* der Entscheidungen kennen und wissen, ob sie sich auf Mathematik, Sprachen, Kunst u. a. beziehen und welche Alternativen es in jeder Entscheidungssituation gibt. Viertens muß man die *Funktion* der Entscheidungen kennen und wissen, ob sie sich auf die Planung, das Programm, die Implementation oder die wiederholte Verwendung des Programms beziehen. Fünftens muß man mit dem *Gegenstand* der Entscheidungen (z. B. Personen, Ereignisse oder Dinge) vertraut sein. Sechstens muß man den *Zeitpunkt* der Entscheidungen genau kennen. Siebentens muß man das Ausmaß an *kritischer Reflektiertheit* der Entscheidungen identifizieren.

Wenn man alle genannten Entscheidungsvariablen berücksichtigt, lassen sich viele verschiedene Entscheidungssituationen im Bildungswesen identifizieren. Deshalb kann man auch mehrere Arten der Evaluation unterscheiden. Aus diesem Grunde sollte man ein Klassifikationssystem für die verschiedenen Arten der pädagogischen Evaluation entwickeln, das die allgemeine konzeptuelle Evaluationsdefinition mit den zahlreichen spezifischen Arten der Evaluation in Beziehung setzt, die sich in einer detaillierten Analyse und Klassifikation pädagogischer Entscheidungssituationen aus einer Berücksichtigung der genannten Variablen gewinnen lassen. Sodann gilt es schließlich, für die identifizierten Klassen pädagogischer Evaluation brauchbare Begriffe zu finden.

Um ein Klassifikationssystem für pädagogische Entscheidungssituationen und Programme zu erarbeiten, konzentrierte ich mich anfangs ausschließlich auf die Funktion der Entscheidungen (Stufflebeam 1967). Meiner Ansicht nach lassen sich die Funktionen von Entscheidungen im Bildungswesen als *Planung*, *Programmgestaltung*, *Implementation* und *modifizierte Programmwiederholung* klassifizieren. *Planungsentscheidungen* richten sich auf erforderliche Reformen und präzisieren ihren Bereich und ihre allgemeinen und spezifischen Ziele. *Programmentscheidungen* richten sich auf die Verfahren, die beteiligten Personen und die zeitlichen und finanziellen Bedingungen für die Implementation der geplanten Akti-

vitäten. *Implementationsentscheidungen* richten sich auf die im Programm intendierten Handlungen. Zu den Entscheidungen, die mit der Frage der *wiederholten, bzw. modifizierten Verwendung des Programms* zusammenhängen, gehören diejenigen über die Beendigung, Weiterführung, Entwicklung oder Veränderung des Programms.

### *Vier Strategien zur Evaluation von Bildungsprogrammen*

In Entsprechung zu diesen vier Arten von Bildungsentscheidungen gibt es auch vier Arten von Evaluation. Sie werden in Tabelle 1 als Kontext-, Input-, Prozeß- und Produktevaluation bezeichnet. *Kontextevaluation* findet während der ersten Phase der Projektplanung statt. *Inputevaluation* erfolgt gleich danach bei der spezifischen Planung des Programms. *Prozeßevaluation* findet während der Implementation des Projekts statt. *Produkt-evaluation* erfolgt im allgemeinen nach der Beendigung des Projekts. Diese vier Formen der Evaluation sollen im folgenden weiter entwickelt werden.

#### Kontextevaluation

Das Hauptziel der Kontextevaluation besteht darin, die Voraussetzungen, unter denen eine Reform erfolgt, und die unbefriedigten Bedürfnisse der Umwelt und die mit ihnen verbundenen Probleme zu bestimmen. Die Umwelt kann z. B. aus den Grundschulen des Zentrums einer großen Stadt bestehen. Die Erforschung dieser Bedingungen könnte ergeben, daß die wirklichen Leseleistungen der Schüler in diesem Schulbezirk weit unter den Erwartungen des Schulsystems liegen. Damit wäre ein Mangel identifiziert, d. h., die Kontextevaluation hätte ergeben, daß die Leseleistungen der Schüler verbessert werden müssen.

Als ersten Schritt in der Kontextevaluation müßten die Schulen die Gründe für die mangelnden Leseleistungen zu erkennen versuchen. Ist der Unterricht der Schüler angemessen? Entspricht das Unterrichtsmaterial ihren Bedürfnissen? Gibt es Sprachbarrieren? Bleiben die Schüler dem Unterricht fern? Sind die Erwartungen der Schulen an diese Schüler erfüllbar? Dies sind meiner Ansicht nach mögliche Probleme und Schwierigkeiten, die verhindern, daß die angestrebten Ziele erreicht werden, und die dadurch dazu führen, daß solche Unzulänglichkeiten entstehen.

Bei der Kontextevaluation beginnt man mit einer konzeptuellen Analyse, um den Untersuchungsbereich mit seinen wichtigsten Teilbereichen zu identifizieren und zu begrenzen. Sodann werden empirische Untersuchungen mit Stichprobenerhebungen, Umfragen und standardisierten Tests durchgeführt. Das Ziel der Kontextevaluation besteht darin, die Diskre-

Tabelle 1: Das CIP-Evaluationsmodell – Ein Klassifikationsschema der Strategien zur Evaluation pädagogischer Reformen (Stufflebeam 1967)

Die Strategien	Kontextevaluation	Inputevaluation	Prozeßevaluation	Produktevaluation
Ziel	Definition des <i>Programmkontexts</i> , Identifikation und Einschätzung der <i>Bedürfnisse</i> in dem Kontext und Identifikation und Beschreibung der <i>Probleme</i> , die mit den Bedürfnissen verbunden sind.	Identifikation und Abschätzung der Systemmöglichkeiten, der verfügbaren <i>Input-Strategien</i> und der <i>Pläne</i> zur Implementation der Strategien.	Identifikation oder Voraussage, der Unzulänglichkeiten des den Prozeß steuernden Plans oder seiner Implementation und die Aufzeichnung von <i>Ergebnissen und Aktivitäten des Prozesses</i> .	In-Beziehung-setzen der <i>Ergebnisse</i> mit den Lernzielen, dem Kontext, dem Input und dem Prozeß.
Methode	Individuelle Beschreibung der wichtigsten Teilsysteme des Kontexts unter relevanten Gesichtspunkten; Vergleich wirklicher und beabsichtigter Inputs und Outputs der Teilsysteme; Analyse möglicher Gründe für die Diskrepanz zwischen Wirklichkeit und Intention.	Beschreibung und Analyse der verfügbaren menschlichen und materiellen Ressourcen, Lösungsstrategien und Verfahrenspläne in bezug auf Relevanz, <i>Durchführbarkeit</i> und Wirtschaftlichkeit während der Durchführung.	Beachtung der möglichen im Prozeß auftretenden Hindernisse, der Aktivitäten und <i>ständige Aufmerksamkeit</i> gegenüber <i>unerwarteten</i> Hindernissen.	Operationale Definition und Messung der mit den Zielen verbundenen Kriterien durch Vergleich der Meßwerte mit im voraus bestimmten Normen und durch Interpretation des Ergebnisses in bezug auf die aufgezeichneten Input- und Prozeßinformationen.
Beziehung zum Fällen von Entscheidungen im Reformprozeß	Entscheidungen über die <i>Ausgangsbedingungen</i> , die <i>Ziele</i> , die zur Verbesserung der Situation dienen sollen, und die <i>Lernziele</i> , die zur Problemlösung, d. h. zur Planung der benötigten Reformen bestimmt sind.	Auswahl der <i>Finanzierungsquellen</i> , der <i>Lösungsstrategien</i> und <i>Verfahrenspläne</i> , d. h. systematische Planung der Reformaktivitäten.	<i>Implementation und Verbesserung des Programmplans</i> und des Verfahrens, um z. B. den Verlauf wirklich zu kontrollieren.	Entscheidung über die <i>Weiterentwicklung</i> , <i>Beendigung</i> , <i>Modifikation</i> oder <i>Schwerpunktverlagerung</i> einer Reformaktivität und Verbindung der Aktivität mit anderen wichtigen Phasen des Reformprozesses, z. B. neu in Erscheinung tretenden Reformaktivitäten.



panzen zwischen intendierten und wirklichen Situationen für alle Teilbereiche des untersuchten Gesamtbereichs darzulegen und somit die benötigten Reformen zu identifizieren. Schließlich gehören empirische und konzeptuelle Analysen, Theorien und Ansichten von Autoritäten zur Kontextevaluation, um die mit dem Mangel und den Unzulänglichkeiten verbundenen Probleme zu beurteilen.

Zu den Entscheidungen, für die die Kontextevaluation Daten zur Verfügung stellen soll, gehören Entscheidungen über die Ziele, die sich aus den Bedürfnissen ergeben und mit deren Hilfe die Probleme gelöst werden sollen. Derartige Entscheidungen werden im allgemeinen in einleitenden Abschnitten von Projektanträgen an Ministerien oder Stiftungen sichtbar.

### Inputevaluation

Um über die Verteilung von Ressourcen zur Realisierung von Programmzielen zu entscheiden, bedarf es einer Inputevaluation. Ihr Ziel besteht darin, die relevanten Möglichkeiten des Antragstellers, die Strategien und Pläne zur Realisierung der Programmziele und der entsprechenden Lernziele zu identifizieren und zu beurteilen. Das Ergebnis einer Inputevaluation ist die Analyse von alternativen Verfahrensplänen im Hinblick auf mögliche Kosten und möglichen Gewinn.

Insbesondere werden die alternativen Pläne unter Bezug auf die folgenden Aspekte beurteilt: Ressourcen, Zeit, eventuelle Unzulänglichkeiten bei den intendierten Realisierungsverfahren, die Möglichkeiten und Kosten ihrer Überwindung, die Relevanz von Plänen im Hinblick auf Programmziele und die Gesamtmöglichkeiten des Plans für die Realisierung der Programmziele. Im allgemeinen liefert Inputevaluation Informationen für eine Entscheidung darüber, ob für die Realisierung der Ziele äußere Hilfe in Anspruch genommen werden soll und welche Strategien dabei gewählt werden müssen, ob also z. B. bereits verfügbaren Bildungsmöglichkeiten oder der Entwicklung neuer Strategien der Vorzug zu geben ist, und welcher Plan oder welche Vorgehensweise für die Implementation der ausgewählten Strategien verwendet werden soll.

Bislang fehlen im Bildungswesen Methoden der Inputevaluation. Zu den bisher verbreiteten Praktiken gehören Kommissionsberatungen, eine Berücksichtigung der Fachliteratur und die Befragung von Experten. In einigen Bereichen gibt es bereits formale Instrumente, um den Entscheidungsträgern bei Inputentscheidungen behilflich zu sein. Bei der Planung von Testprogrammen kann man in *Buros Mental Measurements Yearbooks* (1949) eine wesentliche Hilfe finden.

Der pädagogische Forscher, der einen experimentellen Versuchsplan

wählen möchte, kann für die Identifikation und Beurteilung alternativer experimenteller Pläne in dem Kapitel über experimentelle Versuchspläne in Gages Handbook of Research on Teaching (1963) erhebliche Hilfen finden. In diesem Kapitel werden für den Forscher, der vor der Entscheidung über einen experimentellen Versuchsplan steht, die relevanten Alternativen experimenteller Forschung ausführlich dargestellt. Alle diese experimentellen Versuchspläne werden in bezug auf die Kriterien innerer und äußerer Validität beurteilt. Sodann werden für alle erwähnten Versuchspläne mögliche Probleme und Schwierigkeiten in der Durchführung angegeben.

Entscheidungen aufgrund von Inputevaluation führen im allgemeinen in den Anträgen an die Ministerien und Stiftungen zu einer Spezifikation der Verfahren, Materialien, Zeitpläne, Stellenanforderungen und des Budgets. Die Anträge werden von den Geldgebern wieder einer Inputevaluation unterzogen, um danach über eine Finanzierung der vorgeschlagenen Projekte zu entscheiden. Stiftungen und Ministerien haben im allgemeinen für ihre Inputevaluation Experten als Berater und Beurteiler.

### Prozeßevaluation

Wenn die Richtung des Vorgehens bestimmt worden ist und die Implementation des Plans begonnen hat, brauchen Projektleiter und die anderen Verantwortlichen eine systematische Prozeßevaluation, um dadurch eine kontinuierliche Kontrolle und Verbesserung der Pläne und Verfahren bewirken zu können. Die Aufgabe der Prozeßevaluation besteht darin, während der einzelnen Stadien der Implementation Unzulänglichkeiten im Verfahrensplan oder in seiner Durchführung zu entdecken oder vorauszusagen. Die Gesamtstrategie zielt darauf, die möglichen Ursachen für Fehlschläge in einem Projekt zu identifizieren: Zu diesen möglichen Ursachen gehören die interpersonellen Beziehungen zwischen den Mitarbeitern, die Kommunikationsstrukturen, das Verständnis und die Unterstützung der Intentionen des Programms durch die Programmentwickler und die Adressaten sowie die Angemessenheit der Ressourcen, der schulischen Anlagen, der zeitlichen Planung und die Eignung der Mitarbeiter.

Im Unterschied zur Evaluation mit einem experimentellen Versuchsplan erfordert Prozeßevaluation weder eine kontrollierte Zuordnung der am Versuch beteiligten Personen in Versuchs- und Kontrollgruppen noch konstant gehaltene Versuchsbedingungen. Ihre Aufgabe besteht darin, den Mitarbeitern des Projekts dabei zu helfen, ihre alltäglichen Entscheidungen ein wenig rationaler zu fällen, um so die Qualität der Programme zu verbessern. In der Prozeßevaluation ist der Evaluator bereit, an dem Pro-

gramm in seiner augenblicklichen und modifizierten Form mitzuarbeiten und die Gesamtsituation so gut wie möglich zu berücksichtigen. Dabei sollte er sich bemühen, bei den wichtigsten Aspekten des Projekts möglichst empfindliche und nicht intervenierende Verfahren der Datensammlung zu verwenden. Eine solche Evaluation ist multivariat; nicht alle wichtigen Variablen lassen sich vor dem Beginn des Projekts spezifizieren. Der Prozeßevaluator konzentriert sich vor allem auf die in der Theorie des Programms entwickelten Variablen, aber er muß auch bereit sein, seine Aufmerksamkeit auf unerwartete, aber wichtige Ereignisse zu richten. Bei einer Prozeßevaluation werden die Informationen täglich gesammelt, systematisch organisiert, periodisch, d. h. z. B. wöchentlich, analysiert und so oft vorgetragen, wie die Projektmitglieder solche Informationen anfordern.

Dadurch erhalten die Entscheidungsträger eines Projekts nicht nur die Informationen, die sie für die Antizipation und Überwindung prozeßbedingter Schwierigkeiten brauchen, sondern auch einen Bericht über den Prozeß der Implementation, der auch für die spätere Interpretation der Projektergebnisse verwendet werden kann.

### Produktevaluation

Produktevaluation dient dazu, nach Beendigung des Projekts seine Wirksamkeit festzustellen. Ihre Aufgabe besteht darin, die Ergebnisse auf die Ziele, den Kontext, den Input und den Prozeß zu beziehen, d. h. die Ergebnisse zu messen und entsprechend zu interpretieren. Dazu muß man die Kriterien, die zu den Intentionen einer Handlung gehören, operational festlegen und messen, die Meßwerte mit den im voraus bestimmten absoluten oder relativen Normen vergleichen und die Ergebnisse mit Hilfe der aufgezeichneten Kontext-, Input- und Prozeßinformationen rational interpretieren. Die Kriterien für die Produktevaluation können entweder primäre oder sekundäre sein (vgl. Scriven 1967). Die sekundären Kriterien beziehen sich auf die Ergebnisse eines Programms, die zur Erreichung der Verhaltensziele beitragen. Clark und Guba haben für pädagogische Innovationen 1965 eine Taxonomie von Zielen mit den dazu gehörenden Kriterien entwickelt, deren Schema ich in Tabelle 2 adaptiert habe. Die primären Kriterien beziehen sich vor allem auf Verhaltensziele, für deren Identifikation Blooms Taxonomy of Educational Objectives (1954) nützlich ist.

Im Reformprozeß bietet die Produktevaluation Informationen, mit deren Hilfe über die Weiterentwicklung, Beendigung, Modifikation einer Reform entschieden und aufgrund derer diese Innovation mit anderen Phasen des Reformprozesses verbunden werden soll. So kann z. B. die Produktevalua-

tion eines Programms, mit dem die Lernbereitschaft von Schülern aus sozio-kulturell benachteiligten Familien angeregt werden sollte, zeigen, daß die Programmziele gut erreicht worden sind, und daß das entwickelte innovative Programm auch auf andere Schulen übertragen werden kann.

Nachdem ich diese vier Formen der Evaluation dargestellt habe, soll die Methodologie ihrer Implementation im nächsten Abschnitt dieses Beitrags entwickelt werden.

### *Die Struktur von Evaluationsplänen*

Wenn ein Evaluator eine Evaluationsstrategie, d. h. z. B. Kontext-, Input-, Prozeß- oder Produktevaluation, gewählt hat, muß er einen Plan (design) für ihre Durchführung auswählen oder entwickeln. Das ist eine schwierige Aufgabe, weil es nur wenige generalisierbare Evaluationspläne gibt, die den Erfordernissen des Bildungswesens genügen. Daher müssen Pädagogen Evaluationspläne im allgemeinen ganz neu entwickeln.

In dem folgenden Abschnitt dieses Beitrags sollen einige allgemeine Richtlinien für die Entwicklung von Evaluationsplänen behandelt werden. Dabei möchte ich die Struktur von Evaluationsplänen im Bildungswesen darzustellen versuchen. Hoffentlich werden diese allgemeinen Ausführungen den Pädagogen bei ihren Versuchen, Evaluationspläne zu entwickeln, behilflich sein, und hoffentlich werden die folgenden Ausführungen Experten in Methodenfragen dazu anregen, generalisierbare Pläne für die Kontext-, Input-, Prozeß- und Produktevaluation zu entwickeln.

### **Definition des Plans**

Im allgemeinen dient ein Plan zur Vorbereitung einiger Entscheidungssituationen, die zur Realisierung bestimmter Ziele führen sollen. Diese Definition besagt dreierlei:

- 1) Man muß die Ziele bestimmen, die durch die Realisierung des Plans erreicht werden sollen. In einer Produktevaluation könnte ein solches Ziel z. B. in der Untersuchung darüber bestehen, ob alle Schüler in einem Leseprogramm bestimmte Leseleistungen und Lesefertigkeiten erreichen.
- 2) Man muß die Entscheidungssituationen während der Realisierung des Evaluationsziels identifizieren. In dem Beispiel vom Leseprogramm müßte man die Meßverfahren bestimmen, die sich für die Einschätzungen der Lesefertigkeit eignen.
- 3) Der Evaluator muß in allen identifizierten Entscheidungssituationen zwischen möglichen Alternativen wählen.

Tabelle 2. Eine Prozeßtafel, die die Rolle der Evaluation im Prozeß der Bildungsreform darstellt. Sie beruht auf D. L. Clark und E. G. Guba (1965) und ist abgedruckt aus: D. L. Stofflebeam (1966)

	Institution	Ziel	Prozeß	Kriterien	Beziehung zur Innovation
<b>F O R S C H U N G</b>	Universitäten, Forschungs- und Entwicklungszentren und Bildungszentren	Verbesserung des Wissens, d. h. darstellen, in Beziehung setzen, konzeptualisieren und überprüfen.		Validität (innere und äußere).	Liefert eine Basis für eine Innovation.
<b>E N T W I C K L U N G</b>	Universitäten, Forschungs- und Entwicklungszentren, Bildungszentren.	<p>Formulieren einer neuen Lösung eines oder mehrerer Probleme, d. h. innovieren.</p> <p>Entwurf eines Plans zur Konstruktion der Innovation.</p> <p>Entwicklung der Komponenten, d. h. Konstruktion.</p> <p>Integration der Komponenten in ein funktionierendes System, z. B. Beendigung der Entwicklung der Innovation für den Verkauf.</p>		<p>Augenscheinvalidität (face validity); geschätzte Funktionsfähigkeit; Einfluß (relativer Beitrag).</p> <p>Durchführbarkeit (Produktion und Benutzung); Handlichkeit (leicht zu benutzen, zu kontrollieren und in der Verwendung zu unterweisen).</p> <p>Spezifikation des Plans; individuelle Verhaltensweisen.</p> <p>Spezifikation des Plans; Präzisierung aller Verhaltensweisen, Funktionsfähigkeit, Leistungsfähigkeit.</p>	<p>Erzeugt die Innovation.</p> <p>Bewirkt, daß die Innovation den Merkmalen der Zielsituation angemessen ist.</p> <p>Schafft die für die Implementation des Plans notwendigen Komponenten.</p> <p>Erzeugt das koordinierte funktionierende System.</p>

I M P L E M E N T A T I O N	Regierung, Universitäten und Bildungszentren.	<p>Schafft verbreitete Kenntnis der Innovation bei den Praktikern, d. h. informiert.</p> <p>Möglichkeit, die Qualität der Reform zu überprüfen und zu beurteilen, d. h. sie überzeugend machen.</p>		<p>Verständlichkeit, Zuverlässigkeit, Überzeugungskraft; Einfluß (Ausmaß, in dem Hauptziele erreicht werden).</p> <p>Glaubwürdigkeit; Angemessenheit; Bewertung.</p>	<p>Informiert über die Innovation.</p> <p>Überzeugt von der Innovation.</p>
A D A P T A T I O N	Universitäten, Bildungszentren und Schulen.	<p>Ausbildung der Lehrer in den Schulbezirken, mit der Innovation umzugehen und sie durchzuführen, d. h. Personalausbildung.</p> <p>Vertrautmachen mit der Innovation und Schaffen einer Grundlage zur Einschätzung der Qualität, des Wertes, der Angemessenheit und der Verwendbarkeit der Innovation in einer bestimmten Institution.</p> <p>Die Charakteristika der Innovation und der sie implementierenden Institution aufeinander beziehen.</p> <p>Die Innovation als eine anerkannte Komponente des Systems assimilieren, d. h. etablieren.</p>		<p>Quantität, Kontinuität, Angemessenheit, Motivation und Tüchtigkeit des ausgebildeten Personals.</p> <p>Anwendbarkeit; Durchführbarkeit; Handlung.</p> <p>Wirksamkeit, Leistungsfähigkeit.</p> <p>Kontinuität, Bewertung; Unterstützung.</p>	<p>Errichtet und erhält die Funktionsfähigkeit für die Durchführung der Innovation.</p> <p>Prüft die Innovation im Kontext einer bestimmten Situation.</p> <p>Operationalisiert die Innovation für die Verwendung in einer bestimmten Institution.</p> <p>Etabliert die Innovation als einen Teil eines bestehenden Programms, überführt sie in eine »Nicht-Innovation«.</p>

Somit würde ein vollständiger Evaluationsplan zahlreiche Entscheidungen über die Durchführung der Evaluation und die Wahl der verwendeten Instrumente enthalten.

Eine Liste mit den in vielen Evaluationsplänen gleichen Entscheidungssituationen, wäre für Evaluatoren außerordentlich nützlich. Sie würde es ihnen ermöglichen, die Probleme des Evaluationsplans systematisch in Angriff zu nehmen. Außerdem könnte sie für die Formulierung der Abschnitte über Evaluation in den Anträgen auf Forschungs- und Entwicklungsprojekte nützlich sein. Die Ministerien und Stiftungen könnten auch mit Hilfe dieser Liste ihre allgemeinen Richtlinien für Anforderungen im Bereich der Evaluation strukturieren. Sie könnte auch zur Bestimmung der Ausbildungsanforderungen wertvoll sein.

In Tabelle 3 wird eine Liste von allgemeinen Entscheidungssituationen für Evaluationspläne zusammengestellt. Sie beruht auf der Voraussetzung, daß die Struktur des Evaluationsplans für die Kontext-, Input-, Prozeß- und Produktevaluation gleich ist. Diese Struktur besteht nach meiner Auffassung aus sechs wichtigen Elementen:

- a) Evaluationsschwerpunkt
- b) Informationssammlung
- c) Informationsorganisation
- d) Informationsanalyse
- e) Informationsbericht
- f) Administration der Evaluation.

Alle sechs Elemente sollen im einzelnen kurz dargestellt werden.

### Evaluationsschwerpunkt

Das erste Element der Struktur eines Evaluationsplans besteht im Evaluationsschwerpunkt. Aus ihm ergeben sich die Ziele der Evaluation und die mit ihrer Realisierung verbundenen Verfahrensfragen. Zu diesem Element des Evaluationsplans gehören vier Aspekte.

Erstens gilt es, die wichtigsten Entscheidungsebenen zu identifizieren, für die Informationen zur Verfügung gestellt werden sollen. So würden z. B. im Titel III des Elementary and Secondary Education Act von den einzelnen Schulen evaluative Informationen für die Ebene des Schulbezirks, des Einzelstaates und des Bundesministeriums benötigt. Bei der Entwicklung eines Evaluationsplans muß man alle relevanten Ebenen berücksichtigen, da man auf den einzelnen Ebenen unterschiedliche Informationen zu verschiedenen Zeitpunkten braucht.

Nachdem die wichtigsten der für die Evaluation relevanten Entscheidungsebenen genannt worden sind, müssen zweitens die Entscheidungs-

situationen auf jeder Ebene identifiziert werden. Bei unseren gegenwärtig geringen Kenntnissen über Entscheidungsprozesse im Bildungswesen liegt darin eine schwierige Aufgabe. Sie zu erfüllen ist jedoch außerordentlich wichtig; sie sollte deshalb sobald als möglich in Angriff genommen werden. Zunächst sollten Entscheidungssituationen im Hinblick auf die relevanten Entscheidungsträger wie Lehrer, Schulleiter, Schulverwaltung und Gesetzgeber bestimmt werden. Sodann sollten die wichtigsten Arten der Entscheidungssituationen, z. B. die Allokation von Mitteln und die Zustimmung zur Programmweiterentwicklung, festgelegt werden. Schließlich sollten diese Arten der Entscheidungssituationen z. B. als Forschung, Entwicklung, Dissemination oder Adaptation klassifiziert werden, wobei dieser Schritt vor allem für die Bestimmung relevanter Evaluationskriterien nützlich ist.

Die identifizierten Entscheidungssituationen sollten dann in bezug auf ihre kritische Reflektiertheit analysiert werden. Relativ weniger wichtige Entscheidungen, für die man die Evaluations-Ressourcen leicht aufbrauchen könnte, sollten nicht berücksichtigt werden. Ferner sollte man abschätzen, wann die ausgewählten Entscheidungssituationen eintreten, so daß man mit Hilfe der Evaluation gewonnene relevante Daten rechtzeitig bereitstellen kann. Schließlich sollte man versuchen, für jede wichtige Entscheidungssituation die Alternativen, die im Verlauf des Entscheidungsprozesses in Frage kommen können, mitzubetrachten.

Wenn die Entscheidungssituationen ausgearbeitet worden sind, müssen drittens die erforderlichen Informationen bestimmt werden. Insbesondere sollte man für jede Entscheidungssituation die Kriterien dadurch festlegen, daß man die Variablen für die Messung und die Normen für die Beurteilung der Alternativen spezifiziert.

Viertens müssen die Richtlinien der Evaluation bestimmt werden. Man muß z. B. entscheiden, ob eine Selbstevaluation oder eine Fremdevaluation erfolgen soll. Ferner gilt es, die Adressaten der Evaluationsberichte zu bestimmen. Schließlich muß man noch das Ausmaß der Datenerhebung durch das Evaluationsteam festsetzen.

### Informationssammlung

Das zweite wichtige Element der Struktur von Evaluationsplänen besteht in der Planung und Sammlung von Informationen. Dieses Element muß in enger Beziehung zu den Kriterien, die im vorigen Abschnitt identifiziert wurden, gesehen werden.

Bei Verwendung dieser Kriterien sollte man zunächst einmal festlegen, welche Informationen gesammelt werden sollen. Dabei muß man vor allem zwei Aspekte berücksichtigen:



- (1) den Ursprung der Informationen, z. B. Schüler, Lehrer, Schulleiter oder Eltern,
- (2) die gegenwärtige Beschaffenheit der Informationen als Ergebnis zufälliger oder systematischer Aufzeichnungen.

Sodann sollte man Instrumente und Methoden zur Sammlung der erforderlichen Informationen, z. B. Leistungstests, Interviews und die relevante Fachliteratur, angeben. Metfessel und Michael (1967) haben eine umfassende Liste von Instrumenten erstellt, die für die Datensammlung im Rahmen der Evaluation relevant sein können.

Für jedes Instrument, das verwendet werden soll, sollte man zunächst das anzuwendende Stichprobenverfahren spezifizieren. Wenn möglich, sollte man einen Schüler nicht zu viele Instrumente bearbeiten lassen. So könnte ein zweckmäßiges Verfahren darin bestehen, eine Stichprobe für das Instrument A zu ziehen, die in dieser Stichprobe enthaltenen Individuen nicht in die Grundgesamtheit zurückzulegen und aus der verbleibenden Gesamtheit eine Stichprobe für das Instrument B zu ziehen usw. In ähnlicher Weise empfiehlt sich, wenn ein Gesamttestwert für den einzelnen Schüler nicht benötigt wird, ein komplexes Stichprobenverfahren, bei dem kein Schüler mehr als eine Stichprobe der Aufgaben eines Tests bearbeitet.

Schließlich sollte man einen Zeitplan für die Datensammlung entwickeln. Er sollte die Beziehungen zwischen der Auswahl der Stichproben und Instrumente und den Terminen für die Informationssammlung im einzelnen festlegen.

### Informationsorganisation

In Evaluationsberichten wird häufig darüber geklagt, daß die Ressourcen nicht ausreichen, alle wichtigen Daten zu verarbeiten. Um diese Situation zu vermeiden, sollte man das dritte Element des Evaluationsplans, die Informationsorganisation, sorgfältig planen. Zur Organisation der Informationen gehört die Entwicklung eines Plans zur Klassifikation der Informationen und zur Bestimmung der Verfahren, der Kodierung des Übertragens auf Lochkarten und des Abrufens von Informationen.

### Informationsanalyse

Das vierte wichtige Element des Evaluationsplans besteht in der Analyse der Informationen. Ihre Aufgabe ist es, für die deskriptive oder statistische Analyse der Informationen zu sorgen, die den Entscheidungsträgern zur Verfügung gestellt werden sollen. Dazu gehören auch Interpretationen und Empfehlungen. Wie bei der Organisation der Informationen muß der Evaluationsplan entsprechende Mittel für die Durchführung dieser Analysen

vorsehen. Diese Aufgabe sollte einem qualifizierten Mitglied des Evaluationsteams oder einem besonderen Team zugeteilt werden, das sich auf die Probleme statistischer Analysen spezialisiert hat. Die für die Analyse der Informationen verantwortlichen Mitarbeiter müssen auch an der Planung der Analyseverfahren beteiligt sein.

### Informationsbericht

Das fünfte Element eines Evaluationsplans bildet der Bericht der Informationen. Er zielt darauf ab, den Entscheidungsträgern die benötigten Informationen rechtzeitig in benutzbarer Form zur Verfügung zu stellen. In Übereinstimmung mit den Grundsätzen und Richtlinien der Evaluation sollten die Adressaten identifiziert werden. Sodann sollten die Verfahren bestimmt werden, um jedem Adressaten die für ihn relevanten Informationen zur Verfügung zu stellen. Ferner muß das Ausmaß und die Form des Evaluationsberichts festgelegt werden.

### Administration der Evaluation

Das letzte Element des Evaluationsplans besteht in der Administration der Evaluation. Ihre Aufgabe liegt darin, die Durchführung des Evaluationsplans zeitlich zu koordinieren. Erstens muß daher ein Gesamtzeitplan für den Ablauf der Evaluation entwickelt werden. Dazu empfehlen sich Verfahren wie die Program Evaluation and Review Technique (PERT). Zweitens muß man die Stellenanforderungen bestimmen. Drittens gilt es, die erforderlichen Mittel festzulegen, um den Grundsätzen und Richtlinien für die Durchführung der Evaluation gerecht zu werden. Viertens muß man untersuchen, inwieweit der Evaluationsplan valide, reliable, zuverlässige, aktuelle und überzeugende Informationen liefern kann. Fünftens gilt es, Verfahren zu entwickeln, um den Evaluationsplan regelmäßig auf den neuesten Stand zu bringen. Sechstens muß man einen Finanzierungsplan für die Evaluation ausarbeiten.

#### Tabelle 3: Entwicklung eines Evaluationsplans

Die logische Struktur eines Evaluationsplans ist für die Kontext-, Input-, Prozeß- oder Produktevaluation gleich. Sie enthält folgende Elemente:

##### *A. Evaluationsschwerpunkt*

1. Identifikation der wichtigsten Entscheidungsebenen (z. B. örtliche, einzelstaatliche und/oder bundesstaatliche)
2. Planung und Beschreibung aller Entscheidungssituationen auf jeder Entschei-

dungsebene in bezug auf ihren Schwerpunkt, die kritische Reflektiertheit, den Zeitpunkt und die Komposition der Alternativen

3. Bestimmung der Kriterien für jede Entscheidungssituation durch Spezifikation der Variablen für die Messungen und der Normen für die Beurteilung von Alternativen
4. Definition der Grundsätze und Richtlinien, innerhalb deren die Evaluation erfolgen soll

#### B. *Informationssammlung*

1. Spezifikation des Ursprungs der zu sammelnden Informationen
2. Bestimmung der Instrumente und Methoden für die Sammlung der erforderlichen Informationen
3. Spezifikation des anzuwendenden Stichprobenverfahrens
4. Spezifikation der Bedingungen und des Zeitplans für die Informationssammlung

#### C. *Informationsorganisation*

1. Erstellung eines Plans für die Informationen, die gesammelt werden sollen
2. Bestimmung der Mittel zur Kodierung, Organisation, Speicherung und zum Wiederabruf der Informationen

#### D. *Informationsanalyse*

1. Auswahl der analytischen Verfahren, die angewendet werden sollen
2. Bestimmung der Mittel zur Durchführung der Analyse
3. Spezifikation des Ausmaßes und der Form der Evaluationsberichte
4. Zeitplan des Informationsberichts

#### E. *Informationsbericht*

1. Definition der Adressatengruppe
2. Bestimmen der Mittel der Informationsvermittlung
3. Festlegen des Formats des Evaluationsberichts
4. Planung der Elemente für die Darstellung der Information

#### F. *Administration der Evaluation*

1. Zusammenfassung des Evaluationsplans
2. Bestimmung der für die Evaluation erforderlichen Mitarbeiterstellen und Finanzen
3. Spezifikation der Mittel, um die Evaluation gemäß ihren Grundsätzen und Richtlinien durchzuführen
4. Evaluation der Möglichkeiten des Evaluationsplans, valide, reliable, zuverlässige, aktuelle und überzeugende Informationen zu liefern
5. Spezifikation und zeitliche Planung der Mittel, um den Evaluationsplan regelmäßig auf den neuesten Stand zu bringen
6. Bereitstellung eines Budgets für das ganze Evaluationsprogramm

Ich bin am Ende meiner Ausführungen angelangt. Obwohl ich nur einen groben Überblick über einige Probleme der Evaluation im Bildungswesen gegeben habe, wird deutlich geworden sein, daß die Planung und Durchführung pädagogischer Evaluation ein höchst komplexes und schwieriges Unternehmen ist. Es bedarf einer erheblichen Anstrengung aller im Bereich pädagogischer Evaluation arbeitenden Wissenschaftler, um entsprechende Fortschritte zu erzielen. Bleiben sie aus, wird meiner Ansicht nach das Erziehungswesen darunter leiden, daß die für wichtige Entscheidungen erforderlichen Informationen fehlen.

*Evaluation* sein. Denn sie hilft weder bei der Datenerhebung noch der Interpretation der Daten. Aufgrund ihrer Zielsetzung enthält sie Kriterien (zur Auswahl der Variablen) und überläßt die Beantwortung der Frage nach Normen (Welche »ratings« sind angemessen?) dem Beobachter.

\* Die vierte Auflage der *Evaluative Criteria* erschien 1969; National Study of Secondary School Evaluation, 1785 Massachusetts Avenue, N. W. Washington D. C. 200 366.

DANIEL L. STUFFLEBEAM: *Evaluation als Entscheidungshilfe*

Übersetzung von Gudrun Eggert und dem Herausgeber.

Originaltitel: *Evaluation as enlightenment for decision making*, in: *Improving educational assessment and an inventory of measures of affective behavior*, hg. W. H. Beatty, Association for Supervision and Curriculum Development, National Education Association, Washington D. C. 1969, 41–73.

1 Public Law 89–10: The Elementary and Secondary Education Act of 1965, Titel I.

2 Diese Kriterien sind auf den Seiten 70–71 der Titel – III – Richtlinien aufgeführt: *A manual for project applicants and grantees*, U.S. Office of Education, Washington D. C. 1967.

3 Egon G. Guba: *Evaluation and the process of change*, Notizen und Arbeitspapiere bezüglich der Administration von Programmen des Titel III des Public Law 89–10, The Elementary and Secondary Education Act of 1965, erweitert durch Public Law 89–750, April 1967, 312.

4 a. a. O.

5 Citizen Committee for Children of New York, Inc., Newsletter, Bericht von Mrs. Nathan W. Levin, Vorsitzende der Educational Services Section, vor dem Subcommittee on the Elementary and Secondary Education Act of the Education and Labor Committee of the House of Representatives, März 1967.

6 Egon G. Guba: *Methodological Strategies for educational change*, ein Referat, das auf der Tagung über Strategien zur pädagogischen Reform gehalten wurde, die vom 8.–10. November 1965 in Washington D. C. stattfand.

7 Das EPIE-Forum. Monatliche Veröffentlichung des Instituts für Informationsaustausch über Erziehungsprodukte, das von und für Pädagogen geschaffen wurde. New York: Educational Products Information Exchange Institute.